

Januar 2021

Matthias Kettner (federführend), Susanne Hahn, Rainer Adolphi

Abschlussbericht der AG Philosophische Digitalisierungsforschung (PhiDiFo 2020)

Philosophische Digitalisierungsforschung – Ansatz und Ziele

Die „AG Philosophische Digitalisierungsforschung“ thematisiert Digitalisierung als eine Fülle von *kulturellen Prozessen* und verfolgt dabei mehrere Ziele: (1) eine Binnenverständigung innerhalb der Philosophie, welche Phänomene der Digitalisierung eine philosophische Perspektive herausfordern und welches die spezifische Leistung der Philosophie ist; (2) eine Lokalisierung, welche Beiträge die Philosophie zu einem interdisziplinären Diskurs leisten kann und welche Beiträge aus anderen Wissenschaften für eine philosophische Auseinandersetzung erforderlich sind; (3) wissenschaftsorganisatorisch – den Aufbau eines Netzwerks von Philosophinnen und Philosophen, die zur Digitalisierung arbeiten (s. Anhang).

Organisiert wurde die AG durch ein Führungsteam (Matthias Kettner, Susanne Hahn, Rainer Adolphi) und assoziierte Mitglieder (Kai Denker, Petra Gehring, Armin Grunwald, Christoph Hubig). Durchgeführt wurden im Jahr 2020 drei zweitägige Workshops, die inhaltlich mit den gewählten Schwerpunkten „Verständigungsverhältnisse“, „Verantwortungsverhältnisse“, „Vernunftverhältnisse“ einer philosophischen Taxonomie folgen. Aus den Vorträgen (s. Anhang) und Diskussionen, die in einer Publikation zugänglich gemacht werden sollen, lassen sich die folgenden Zwischenergebnisse festhalten:

Eine philosophische Perspektive

Die spezifische Leistung der Philosophie besteht in der systematischen Arbeit an normativen und evaluativen Fragen, in der Identifikation und der Bearbeitung begrifflicher Probleme und in der (Re-)Konstruktion systematischer Zusammenhänge, die eine Einbettung digitalkultureller Phänomene erlauben. Der gewählte Rahmen – Digitalisierung als Fülle kultureller Prozesse – vermeidet eine technozentrische Engführung und erlaubt einen kultur-, sozial-, technik- und wissenschaftsphilosophischen Zugriff.

Nicht auf der informationstechnischen Ebene der digitalen Basistechnologie selbst, sondern erst in unseren Handlungs- und Erfahrungspraktiken entstehen mit Sinn und Bedeutung erfüllte kulturelle Formen. Diese Handlungs- und Erfahrungspraktiken umfassen sämtliche Praktiken, die unsere kulturelle Lebenswelt ausmachen. Uns interessiert aber nicht nur, wie fortschreitende Digitalisierung sich de facto in Veränderungen der Lebenswelt zeigt, sondern überdies, welche Veränderungen der Handlungsmacht von Menschen und nichtpersonalen „Agenten“ damit einhergehen und wie Spielräume möglicher Lebensweisen sich dadurch zum Besseren oder Schlechteren hin verschieben – eine Perspektive der Ethik.

Ergebnisse (1): Der Problemkomplex *Opazität*

Die Durchdringung der Lebenswelt mit neuen Digitaltechnologien, vor allem solchen, die sich dem Einsatz maschinellen Lernens verdanken wie z.B. in der Sprach- und Bilderkennung, führt zu einschneidenden Veränderungen des Kommunizierens, des Handelns und der Wissenserzeugung. Der (nicht nur unter GAFAM-CEOs verbreitete) Glaube, dass wirklicher Fortschritt eigentlich nur technischer Fortschritt sein könne, verdrängt die politisch-ethische Frage, wer mit welchem Recht

welche Standards zur Bewertung von Veränderungen als spezifische Fort- oder Rückschritte formulieren dürfen sollte.

Bei der fälligen Kritik am Fortschrittsreduktionismus ist die Eigenart der neuen Digitaltechnologie selbst in Betracht zu ziehen. Durch die mangelnde Nachvollziehbarkeit von Verfahren maschinellen Lernens ergeben sich dramatische Lücken in der Zuschreibung von Verantwortung auf hybride, durch Mensch-Maschine-Schnittstellen vermittelte Aktionsträgerschaften (z.B. algorithmische Entscheidungssysteme in allen möglichen Anwendungsbereichen, von Militär bis Medizin). Die wissenschaftsphilosophisch bedeutsame Mustererkennung aus Daten und ihre Verarbeitung zu prädiktiven Modellen ist aufgrund der Intransparenz einer argumentativen Durchdringung entzogen. Da ihre Anwendung inzwischen auch dort stattfindet, wo die Verschlechterung menschlicher Lebenschancen auf dem Spiel steht, etwa im Bereich des „predictive policing“ oder im Strafvollzug (z.B. „criminal risk assessment algorithms“), ist die wissenschaftsphilosophische und ethische Forderung nach Offenlegung der Prinzipien der Theoriebildung resp. Mustererstellung angezeigt. Derzeitige Forschungsbemühungen im Rahmen von „explainable AI“ sollten sich diesem Desiderat stellen.

Ergebnisse (2): Der Problemkomplex *Kulturelle Übersetzung*

Ein weiterer Problemkomplex ist mit dem Stichwort der *kulturellen Übersetzung* aufgerufen. Hier besteht eine philosophische Aufgabe in der begrifflichen Begleitung und kritischen Reflexion von Denkweisen und Modellen. Ein intentionalistisches Vokabular der Selbstinterpretation, wozu Begriffe wie Person, Akteur, Entscheidung, Freiheit, Gefühl, Wille, Wissen, Erkennen, Selbstbestimmung („Autonomie“), Verantwortung, Vernunft gehören, verändert sich in den vielfältigen Übersetzungen dieses Vokabulars in die Fachsprachen zur Beschreibung technokultureller Systeme, etwa der informationswissenschaftlichen Beschreibung „künstlicher“ Intelligenz oder der sozialkybernetischen Beschreibung „smarter“ Städte. Kulturelle Übersetzung ist performativ: Sinnveränderungen wirken auch zurück. Hier sehen wir eine weitere genuin philosophische Aufgabe: relevante Bedeutungsveränderungen zu beschreiben, einzuordnen und zu bewerten. Genau untersucht werden muss, was passiert, wenn Personen sich selbst in Begriffen von Rechnern zu deuten beginnen, oder wie die Ansprache von Maschinen als Agenten oder Entscheider sich auf vital wichtige normative Ordnungen, vor allem die Rechtskultur, auswirkt. Philosophie sollte hier zur öffentlichen Diskussion durch aktive kulturelle Übersetzung beitragen, indem sie Reizwörter, die als aus den Fachkulturen entlassene Begriffe semantisch verwildern, politisch aber umso folgenreicher werden (wie Algorithmen, AI, Big Data, Cyberwar, Cybercrime, Darknet, Automated Decision Making u.a.m.) sokratisch angreift, sie entzaubert, in Gegenbewegungen dort Komplexität erhöht, wo zu simpel gedacht wird, und an Grundsätzliches erinnert, wo Fachdiskurse zu einer spezialistischen Überkomplexität neigen.

In den Debatten um grundsätzliche Fragen zum Verständnis künstlicher Intelligenz findet sich der Beitrag der Philosophie im Rückgriff auf erstaunlich langlebige und produktiv bleibende gedankenexperimentelle Theorie-Kontroversen zur Frage, ob lernfähige Technofakte ein Verständnis ihrer selbst innerhalb ihrer Welt entwickeln können. Turings „Test“ auf der affirmativen Seite und Searles „Chinese Room“ auf der skeptischen Seite stellen immer noch Referenzpunkte dieser Debatte dar (vgl. <https://plato.stanford.edu/entries/artificial-intelligence/#PhilAI>).

Ergebnisse (3): Der Problemkomplex *Ab-, Ein- und Anspannung der Ethik*

Technische Fortschritte des maschinellen Lernens fordern mehr noch die praktische Philosophie heraus, denn sie berühren ein lebensweltliches Grundverständnis, das die philosophische Ethik umsichtig durchdacht hat: Dass Akteure, wenn sie menschlichen Kommunikationsgemeinschaften angehören, zu solchen werden, die sich untereinander Verantwortung zuschreiben und nach Maßgabe guter Gründe Rechenschaft verlangen und geben können. Die Einbindung von Verfahren maschinellen

Lernens in Handlungs- und Entscheidungsprozesse fordert die etablierten Bedingungen korrekter Zuschreibung von Verantwortung heraus. Ethische Analyse ergibt zum einen, dass *Kontrollierbarkeit* von Handlungen als Kriterium für Verantwortungszuschreibung die *Nachvollziehbarkeit* algorithmischer Modelle voraussetzt. Bedenken angesichts wachsender Möglichkeiten technischer Verhaltenskontrolle erfordern über „explainability“ hinaus *Rechtfertigbarkeit*; „trustworthy AI“ bleibt ein offenes Problem.

Zum anderen wird deutlich, dass der Einsatz einer Technologie, die unsere gut etablierten Verantwortungspraktiken herausfordert, eine Debatte über die dadurch tangierten vitalen Interessen von Individuen notwendig macht: Ist der Einsatz beispielsweise in der medizinischen Diagnostik so fruchtbar und im Vergleich zu menschlicher Fehleranfälligkeit so nützlich, dass massive Bedenken wegen der Nicht-Nachvollziehbarkeit technisch assistierter Entscheidung klar aufgewogen werden?

Ethische Analysen dieser Art, die den anspruchsvollen Reflexionscharakter der Ethik herausstellen, finden jedoch in der Praxis von Politikberatungskommissionen wenig Raum. Vielmehr ist bei vielen Einrichtungen eine Verkürzung des Ethikbegriffs auf Werte und die Trivialisierung von Werttheorie auf Platitüden wie die folgende erkennbar: „AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity“ (<https://futureoflife.org/>). Die Anspannung von Ethik als Problematisierung wird häufig geringer geschätzt als ihr Wert für Zwecke der Entkomplizierung, Akzeptanzbeschaffung, Beruhigung – Zwecke der abspannenden Problementsorgung innerhalb der fachlichen, politischen und öffentlichen Diskurse.

Ein Desiderat künftigen philosophischen Engagements ist die Kommunikation philosophischer Fragestellungen und Herangehensweisen, die versucht, zu aufgeklärter eigener Urteilsbildung in kontroversen Fragen der Digitalisierung öffentlichkeitswirksam beizutragen. Die Vernetzung und die Workshops der AG haben bereits einige wertvolle Erträge hierfür erbracht.

ANHANG: Personen und Themen der AG PhiDiFo 2020